

Weekly Report

December 24, 2017

1 Work

1.1 降维

论文方面已经完成了Introduction、Method、部分Experiments。下周准备将试验数据跑一跑，把结果放上去。

数据又增加了CIFAR10和CIFAR100，我们可能还需要一些规模更大的数据。imagenet的数据由于数据量太大内存太小解压不开，下周使用楼下服务器解压一下。目前已有数据如下：

Table 1: Summary of data sets

| Data set | Size | Dimension | Class number |
|---------------|-----------|-----------|--------------|
| MNIST | 70,000 | 784 | 10 |
| Fashion-MNIST | 70,000 | 784 | 10 |
| CIFAR-10 | 60,000 | 1024 | 10 |
| CIFAR-100 | 60,000 | 1024 | 100 |
| SVHN | 630,420 | 256 | 10 |
| Twitter | 1,193,514 | 200 | None |

目前部分数据上的运行结果：

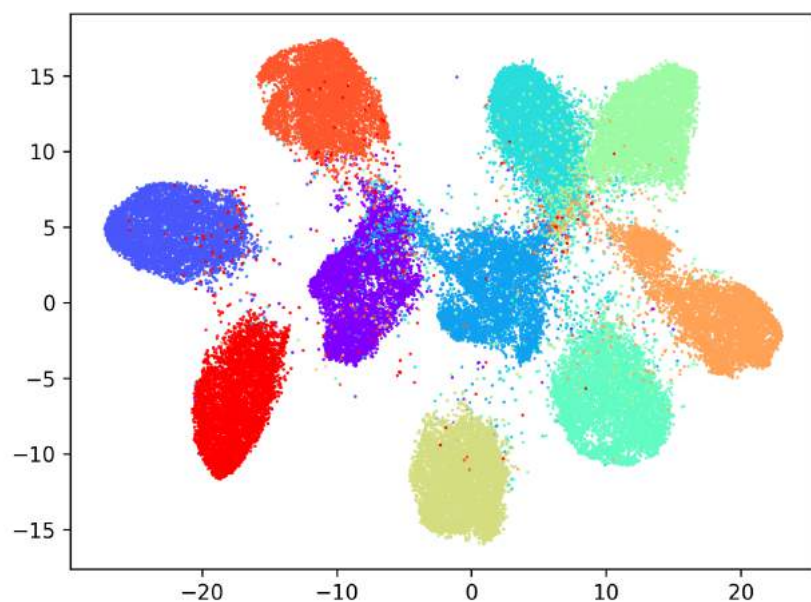


Figure 1: CIFAR100结果（经过训练）。可能是由于把原始的100类绘制的时候只用了10个颜色，所以看起来很奇怪。

Table 2: 工作进度

| TASK | PROGRESS | DATE |
|---------------------|---------------|-------|
| dimension reduction | 1)节省内存2) 调整参数 | 12.30 |
| location2vec专利 | | |
| *2Vec survey | | 1.30 |

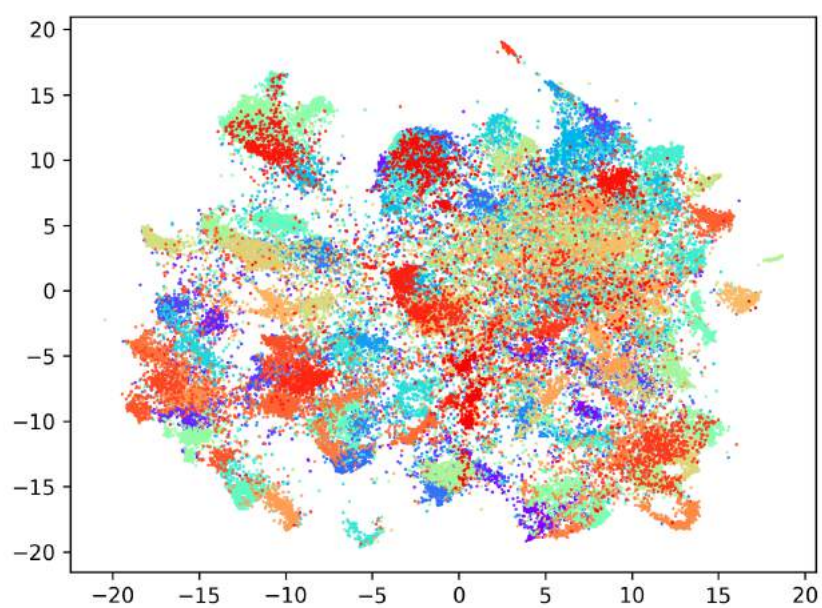


Figure 2: CIFAR100结果（经过训练）。可能是由于把原始的100类绘制的时候只用了10个颜色，所以看起来很奇怪。

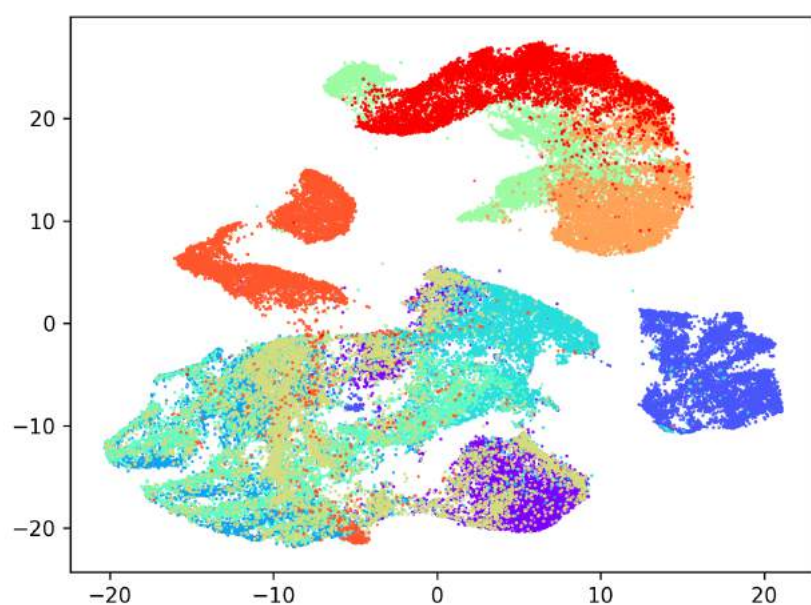


Figure 3: Fashion-MNIST（未经过训练）的结果好像就是这样，原始的数据直接投影就会有几个类重叠在一起。

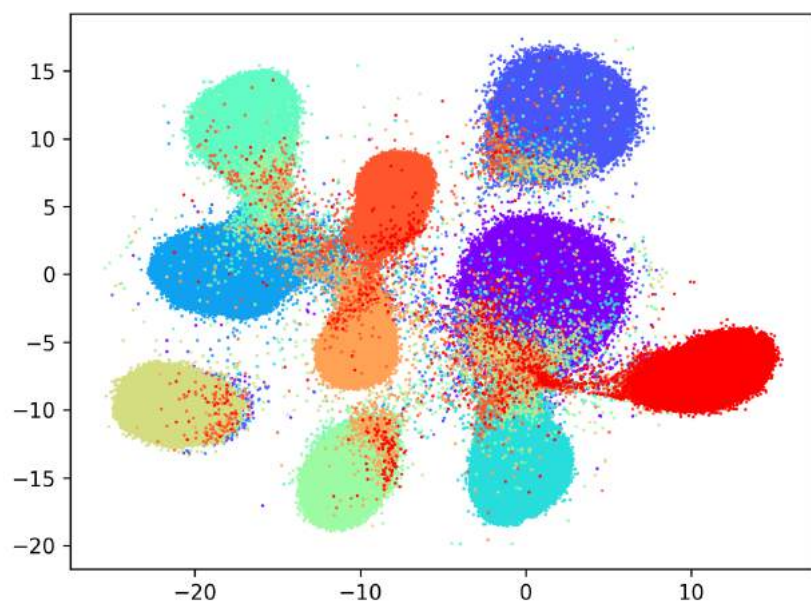


Figure 4: SVHN数据（经过训练）

1.2 工作进度

2 Paper Reading

2.1 Sequence Synopsis: Optimize Visual Summary of Temporal Event Data

文章采用minimum description length方法从时序事件中抽取出频繁的Pattern并进行可视化。

2.2 Understanding a sequence of sequences: Visual exploration of categorical states in lake sediment cores

文章的目标是比较两条时序数据是否相似，这对地理上地层的形成非常有帮助。文章的主要贡献点是提出了可视分析的一套流程，包括识别特殊的序列、聚合相似的特殊序列，检查相关序列组。

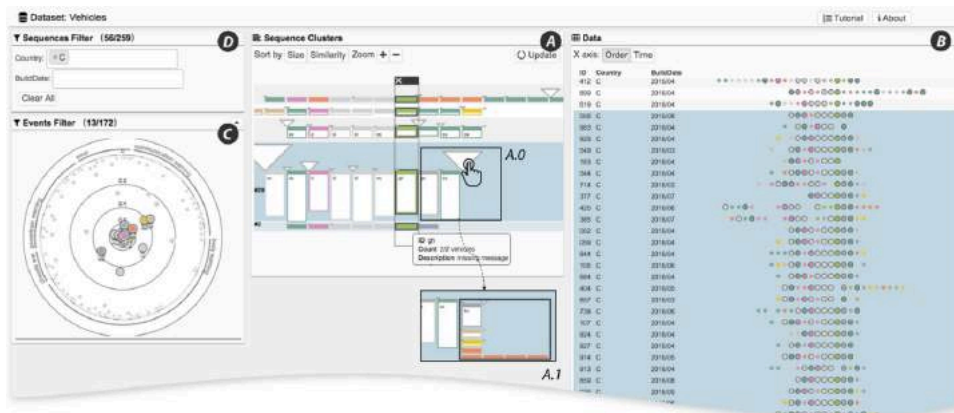


Figure 5: Sequence Synopsis

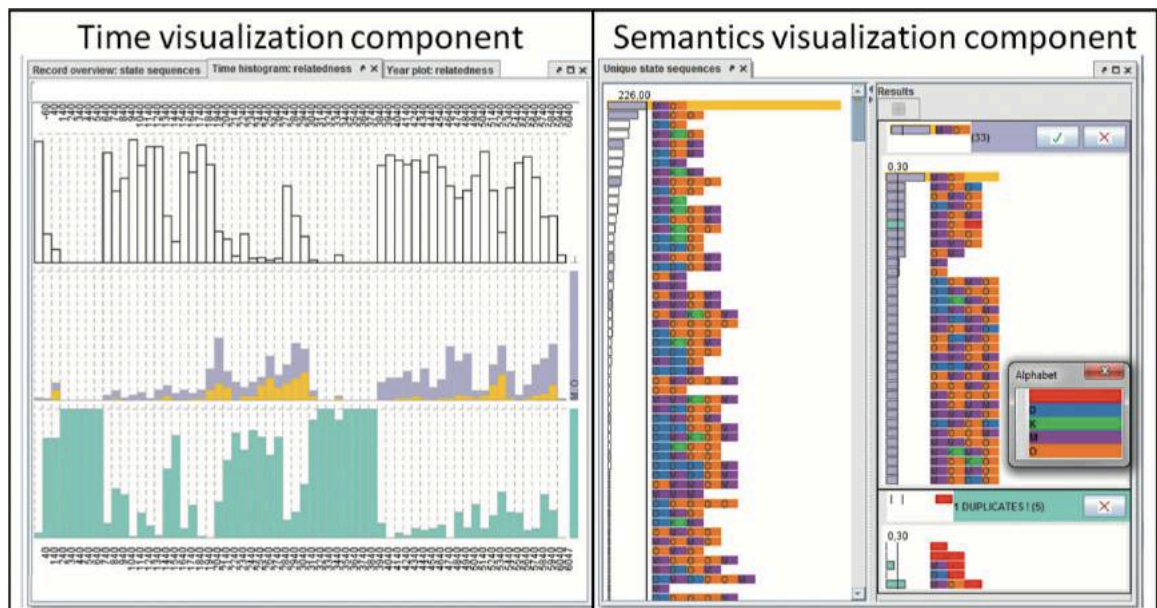


Figure 6: Understanding a sequence of sequences

2.3 VIGOR: Interactive Visual Exploration of Graph Query Results

文章的总体内容是对Graph做查询，并且可视化查询出来的结果。查询的输入是类似于绘制一个子图，比如调研在VAST和KDD上的共同作者，然后把查询出来的节点（作者）用降维的方法绘制在二维上分析。正在思考是否可以用word2vec查询更有意思的内容。

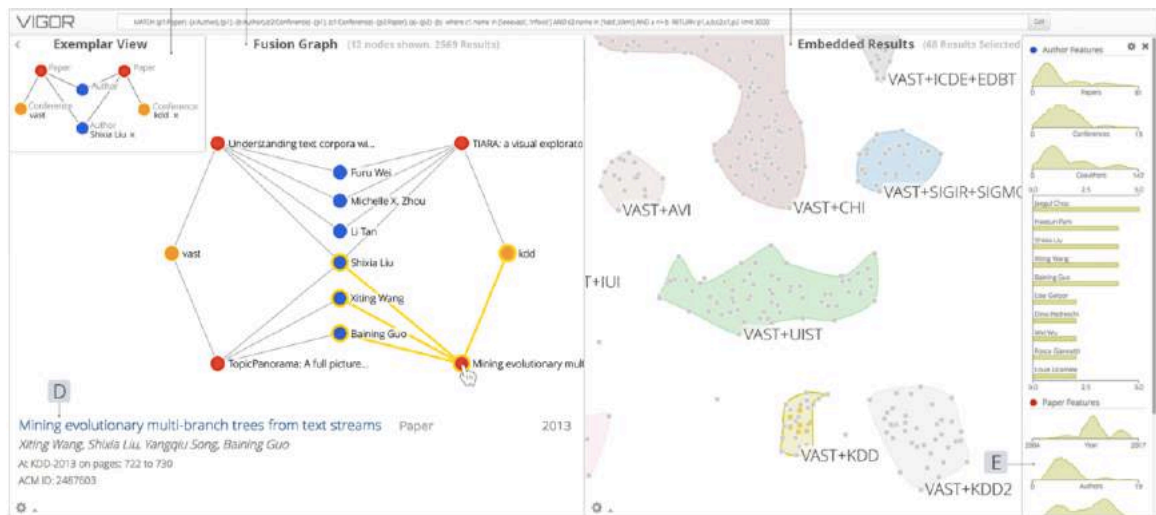


Figure 7: VIGOR